Governance is the ignition, not the handbrake

iManage
Making knowledge work

Contents

Section 1
What agentic AI requires from the systems built to govern it **3**

Section 2
Why governance determines whether context compounds **8**

Section 3
Who's accountable when it goes wrong **10**

Section 4
The MCP protocol connects. An MCP server can govern. That governance doesn't travel. **12**

Section 5
What readiness looks like **15**

Section 6
Governance was never the handbrake **17**

Section 1:

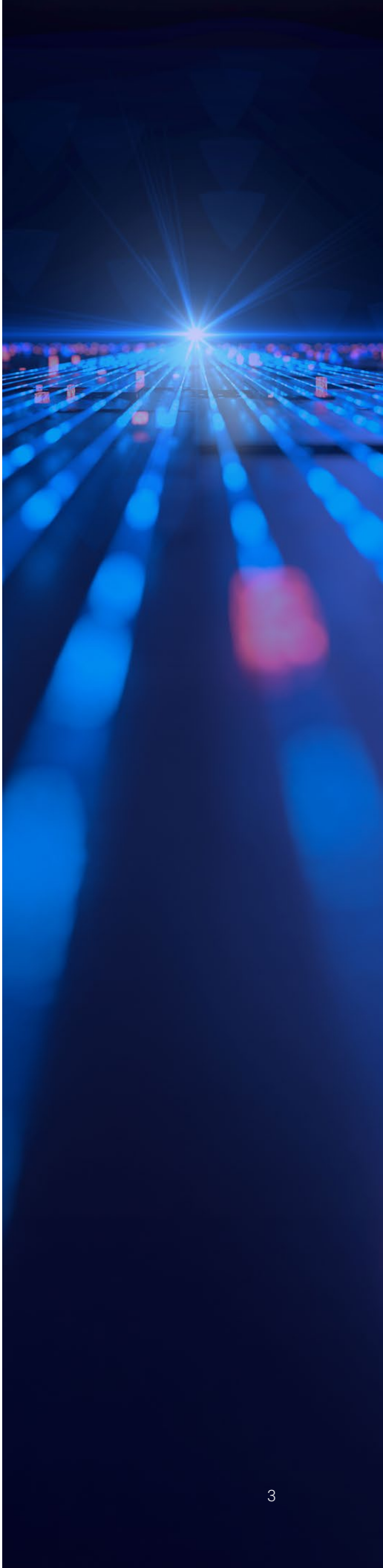
What agentic AI requires from the systems built to govern it

The gap agents don't inherit

Whether it's said or not, every organization adopting AI is relying on judgment that once lived in a person's head — a lawyer who knew which documents were stale or an IT administrator who noticed when access looked wrong. A knowledge worker who filed something in the right place without being told why it mattered.

This judgment was never recorded, because it never had to be. As long as the subject matter expert, a human, was the one acting on it, it was considered "safe."

An AI agent doesn't inherit what was never made explicit. It only knows what's been built into its systems, what someone "tells" it directly, or — absent both — whatever it guesses from patterns that may have nothing to do with your organization. That gap is why governance becomes more than a routine compliance exercise the moment agents start acting on behalf of an organization, and it's the common thread running through everything that follows.

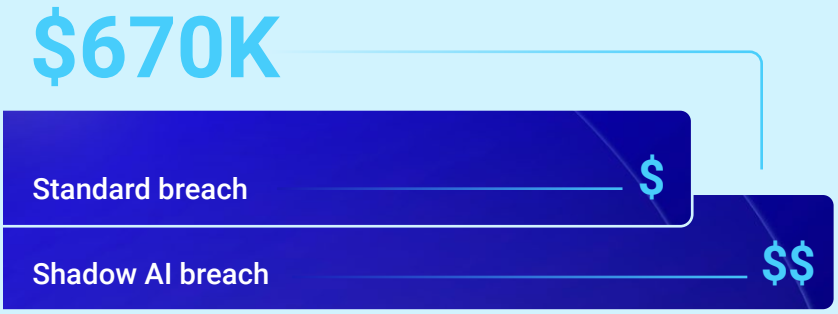


Why the pressure is building now

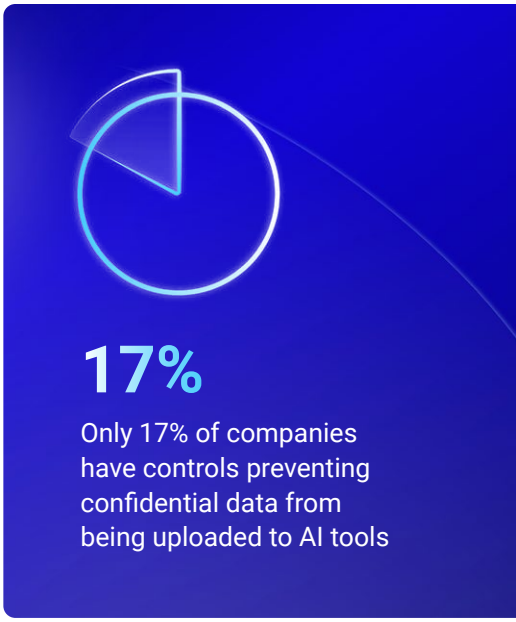
Data volumes continue to increase rapidly. Worldwide data production is expected to rise 22 percent this year alone.[1] But the more pressing reality isn't the data volume itself — it's how much of it agents now access and act on, faster than oversight can track.

Agent use inside large enterprises is projected to grow tenfold by 2027, with the number of calls those agents make rising a thousandfold.[2] Most organizations still can't see where that activity is happening: only 17 percent have controls preventing confidential data from being uploaded to AI tools, and 86 percent lack visibility into their own AI data flows.[3]

Shadow AI, tools used without approval or visibility, already accounts for a fifth of all breaches, and those breaches cost roughly $670,000 more on average than a standard one.[4] In a study of over a thousand office professionals, 88 percent had shared work-related information with a public AI tool, and two-thirds had used AI at work *despite believing their own company policy didn't allow it.*[5]



A Shadow AI breach is $670,000 more costly on average



Closing these gaps isn't just a matter of better tools. Ninety-three percent of firms point to people and process, not the AI itself, as the biggest barrier.[6] And the regulatory picture isn't settling down to make the rest easier: 2026's outlook describes deepening complexity as jurisdictions diverge further, not converge.[7]

With adoption outpacing oversight, vibe-lawyering, the legal profession's version of vibe-coding, becomes a real risk rather than a hypothetical one.

What vibe-lawyering puts at risk

Vibe-lawyering is what happens when a competent lawyer prompts an AI tool for a finished work product and accepts what comes back without the scrutiny they'd normally apply to a first draft from a legal intern. But there's a second, riskier version of it, too: a lawyer working outside their own practice area, using AI to produce something they wouldn't have the expertise to catch an error in *even if they slowed down and reviewed every line.* The American Bar Association has already weighed in directly on this, and its guidance is unambiguous: a lawyer remains responsible for the accuracy of anything filed under their name, regardless of what produced it.[8]

While that responsibility hasn't changed, what has is how easy it's become to quietly skip the step where an expert checks the work (or lacks the expertise to check it meaningfully in the first place) because the output now arrives looking finished.

Even careful lawyers were, until recently, relying on assumptions that were once safe to leave unstated: that someone had already vetted the source material, that stale documents wouldn't surface, and that a sensitive matter was not visible to the wrong eyes. Agentic AI only knows what's been made explicit and doesn't carry any of those assumptions forward.

That is the real reframe — implicit expert human judgment was a working safeguard for a very long time. It stops being one the moment an agent, not a person, is the one making the call.

What “governed” requires

Knowing what data is exposed to AI, having a framework for what’s permitted, and being able to produce a defensible record are the right starting point. Most organizations still can’t confidently claim all three. But a starting point is not the same as a complete one, and treating it as complete is where organizations can stumble later. At minimum, a truly governed environment also has to account for:

- **Monitoring, not only auditing.** Knowing what happened after the fact is not the same as watching it happen in real time, and the difference stops being academic the moment an agent, not a person, is in the driver’s seat.
- **Data quality and currency.** Governing access to data that’s outdated or wrong doesn’t fix the accuracy problem underneath it. Permission isn’t the same as reliability.
- **Ownership and accountability.** Someone needs to be accountable when something goes wrong, and that has to be decided in advance, not determined afterward.
- **Retention and disposition**. Data that should have expired and wasn’t disposed of is still exposed to whatever an agent can reach.
- **The legal right to use the data at all.** Being permitted internally is a different question from having the underlying right, contractual, privacy, or otherwise, to expose that data to AI in the first place.

That gap between auditable and monitored stops being theoretical once agents are acting on their own. Most organizations still treat a human’s actions and an agent acting on that human’s behalf as a single signal, which means an agent quietly accessing documents outside its normal pattern doesn’t register as unusual, because there’s no separate baseline it could violate.

Closing that gap means treating agent activity as its own category, something watched as it happens rather than only reconstructed afterward, while the human who set the agent in motion stays accountable for what it does.

Trusting an agent with a task

A panel of legal and technology leaders converged on five criteria for deciding when an agent can be trusted with a given task:[9]

1. **Define the boundaries** of data access for that specific task.
2. **Ensure reversibility,** so a destructive or irreversible action isn't the first mistake an organization discovers.
3. **Maintain observability,** so the reasoning behind an action isn't a black box.
4. **Gather real evidence of reliability** over repeated tasks, not a one-time impression.
5. **Enforce restrictions from outside the agent,** not only from within its own instructions.

These five criteria don't call for a new governance philosophy, only for existing judgment to be made explicit enough for a system to follow it. An agent can't act on a boundary, threshold, or escalation path that isn't clearly defined. That's the real shift: governance must produce something a machine can execute, not only something a person can be trained on.

Section 2:

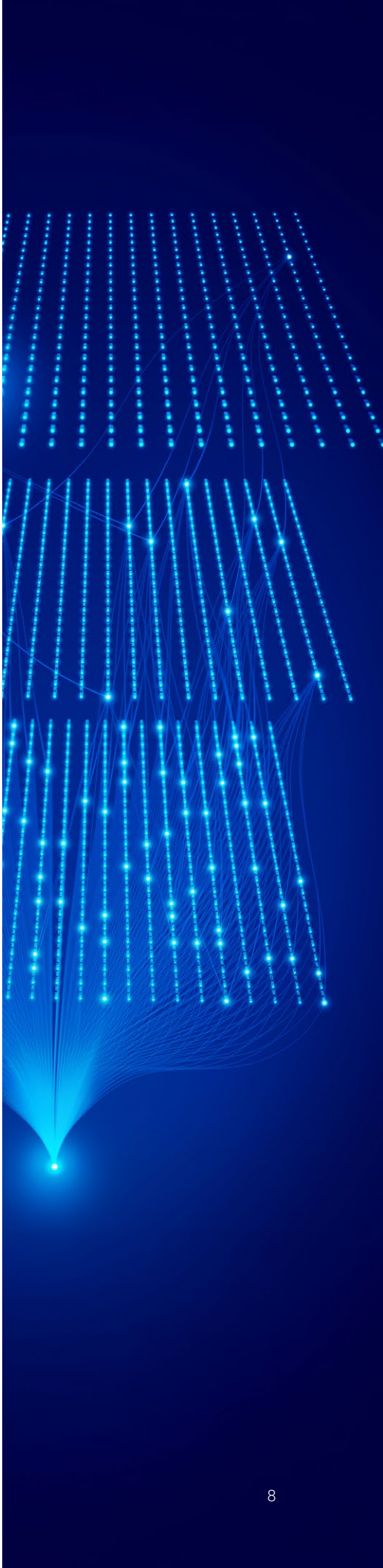
Why governance determines whether context compounds

Everything above is about keeping AI from doing damage. This section is about the upside: governed data doesn't just reduce risk, it becomes a real asset, but only when it carries more than the words in the document.

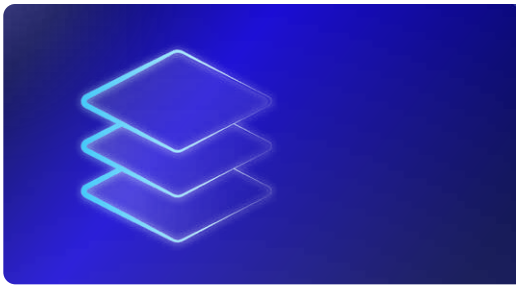
A document sitting on its own is just text. The same document, filed against a matter, inherits metadata: a client, a timeline, a practice area, a team. That single act of filing is what turns a document into something an agent can reason about rather than just read.

The same is true of a few other things most organizations already have and don't think of as valuable:

- **How a document got to its current version.** Documents rarely arrive in final form. The sequence of drafts, and the notes explaining why something changed, is a record of the reasoning behind the work, not just its outcome. This information disappears the moment only the final version is kept.
- **When work happened.** A matter that goes quiet for six weeks and then has forty actions in two days is telling you something a static document never could. An agent reasoning over content alone has no way to know a client went silent or a deadline moved.



- **What scope and staffing looked like.** Who was staffed, what the budget was, how the matter was billed, none of that lives in the document body, because it was never the document's job to hold it. It's the difference between an agent that can say "this clause is standard for a deal this size" and one that has no idea what "this size" even means.



None of that context has to be captured through extra effort. When an AI-suggested tag is accepted by a human, or a document is classified by a system rather than by hand, structure is added to content that never had any. Each classification that follows is faster and more accurate because there is more to draw from.

The important caveat: context only compounds if what's being kept is still useful. Redundant, obsolete, and outdated content doesn't just sit there harmlessly, it dilutes everything around it. An agent reasoning over a corpus full of stale, duplicate, or irrelevant material gets less accurate, not more, no matter how much good context surrounds that clutter. Disposition, deciding what to get rid of, is what protects the value of what's kept. It's more than administrative housekeeping, it's what keeps the context worth trusting.

Governed, well-disposed data is what makes that compounding value durable instead of diluted. Skip the discipline, and the same volume of content works against you instead of for you.

Section 3:

Who's accountable when it goes wrong

As AI takes on more of the work, three questions get harder to leave unanswered:

1. Who is accountable for AI-generated output?
2. Where is human validation required?
3. How is risk weighed against speed?

Legal leaders at a recent industry conference kept arriving at the same idea from different directions: using AI to absorb predictable, repetitive work isn't a threat to the profession, it's a return to what drew people to it. As one general counsel said, the job was never about redlining contracts: it was about judgment, about protecting the organization, about being part of the decision rather than cleaning up after one made without legal input. Another leader described it as a return to an apprenticeship model, teaching junior lawyers how to think rather than where to click.

“There will always be a human element,” one attorney put it, “even if it’s just the moment someone signs their name to the work.” Accountability has to hold that line. AI can assist, but it cannot be the one standing behind the outcome.

Expertise stays firmly in the loop here.

A lawyer’s duty of competence has always meant standing behind the work product, and that doesn’t change because an agent produced the draft. What changes is what competence requires: choosing tools carefully, reviewing output with real scrutiny, and keeping a defensible record of what the AI did and why. Agentic work should be identifiable as agent-led, not indistinguishable from a person’s.

Case in point:

Air Canada

A 2024 case involving Air Canada’s customer-service chatbot is a preview of what happens when that accountability is left ambiguous: a customer relied on the chatbot for guidance on bereavement fares, and its answer didn’t match the airline’s actual policy. British Columbia’s Civil Resolution Tribunal held Air Canada responsible for what its own chatbot told the customer, rejecting the airline’s argument that the bot should be treated as a separate, independently accountable actor.[10]

No organization gets to outsource responsibility to its own tool, and the inherent risks will only multiply as more businesses deploy agents of their own.

Section 4:

- The MCP protocol connects.
- An MCP server can govern.
- That governance doesn't travel.

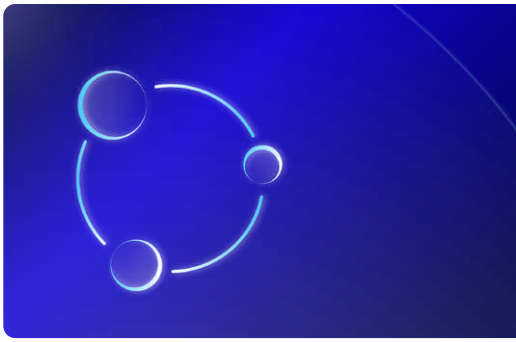
AI tools need to access content, work with it, and sometimes store copies of it. A document gets uploaded from the knowledge work platform into a chat interface, a bulk pull happens for analysis, or a tool ingests and retains content from the platform in its own vault. The governance that protected that content doesn't travel with it. The receiving system doesn't inherit the knowledge work platform's governance. It applies its own.

Each time that content moves, a new version exists outside the platform, disconnected from the one place it was meant to be authoritative. Multiply that across every upload, every pull, every tool, and the idea of a single source of truth diminishes over time. That's a data management problem as much as a governance one: content sitting outside the governed foundation is content the organization may not be actively managing. It just accumulates, ungoverned, in places the organization didn't choose and may not know about.

Model Context Protocol (MCP) is a connection standard: an open protocol that lets any AI system connect to external content through one common interface, replacing the bespoke integration otherwise needed per vendor.

One thing that connection makes easier is reaching data at the source instead of duplicating it at scale first. Previously, exploratory AI analysis over an entire knowledge work platform meant bulk-copying it into one or two separate systems just to get started. With MCP, an AI system can query the platform directly instead, leaving the underlying content in place. But what MCP changes is how easy that is, not what's guaranteed. That same connection can just as easily be used to pull large quantities of data, systematically, at scale, since nothing in the protocol distinguishes a narrow query from a bulk one. That choice belongs to the AI tool or system accessing content in the knowledge work platform. And regardless of how broad or narrow, queries still put something into the requesting system's context window, whether that's an excerpt, a field, a snippet, or metadata like a document's author, date, or matter association.

Governance, though, is a broader question than just access. An MCP server can be implemented to call into the platform's own governed APIs using the same permissions as the requesting user, which means it inherits what those APIs already enforce: for example, permissions scoped to that user, information barriers, or a log of every action. An integration built differently might not provide such governance capabilities. The point is that neither the protocol nor an MCP server inherently provides governance. It's how the MCP server is implemented that determines it. And regardless, once content is pulled through the server into a separate AI tool, that system now owns how it handles governance from there, not the knowledge work platform it was taken from.



To manage and govern content well in light of all this, a shared approach across every party involved is needed: one where everyone understands what happens to content whether it's accessed in place or moves outside the platform, since that movement is what erodes a trusted single source of truth and presents governance considerations.

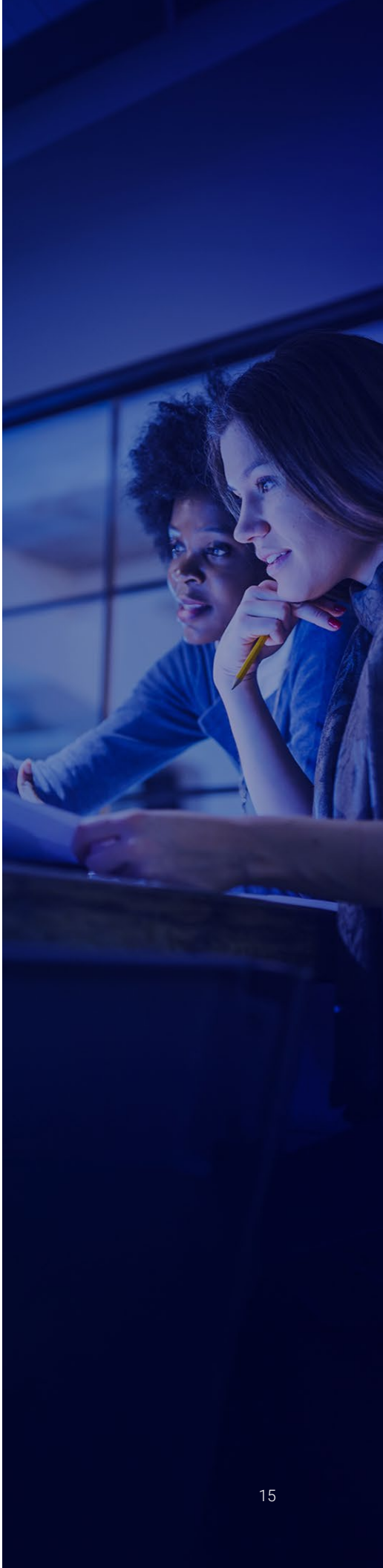
Section 5:

What readiness looks like

Three questions, answered honestly, tell an organization a great deal about whether it's ready to hand real work to an agent. Not a "one and done" exercise, these questions should be answered whenever the scope of what can be delegated to an agent grows.

1 **Do you know what data you have, where it lives, and what AI can see?**

This is the starting question, and most organizations answer it less confidently than they'd like. It means knowing which systems hold the content, which of those an agent can reach, and what's already been done to it before an agent ever touches it.



2 **Is your security model built for agents, not only humans?**

A security model that worked well enough when humans were the only ones browsing it becomes a real liability the moment agents can traverse it without judgment. Although a person could technically open every file they had access to, in practice, they mostly didn't. An AI agent has no equivalent instinct.

3 **Can you prove, defensibly, what your AI did and why?**

Knowing what happened isn't a compliance afterthought, but it isn't the whole answer either: a knowledge worker still has to apply their own scrutiny before putting their name on it.

Readiness compounds the same way governed data does: the organizations that can answer these questions with confidence today will still be equipped to answer them with confidence years from now.

Section 6:

Governance was never the handbrake

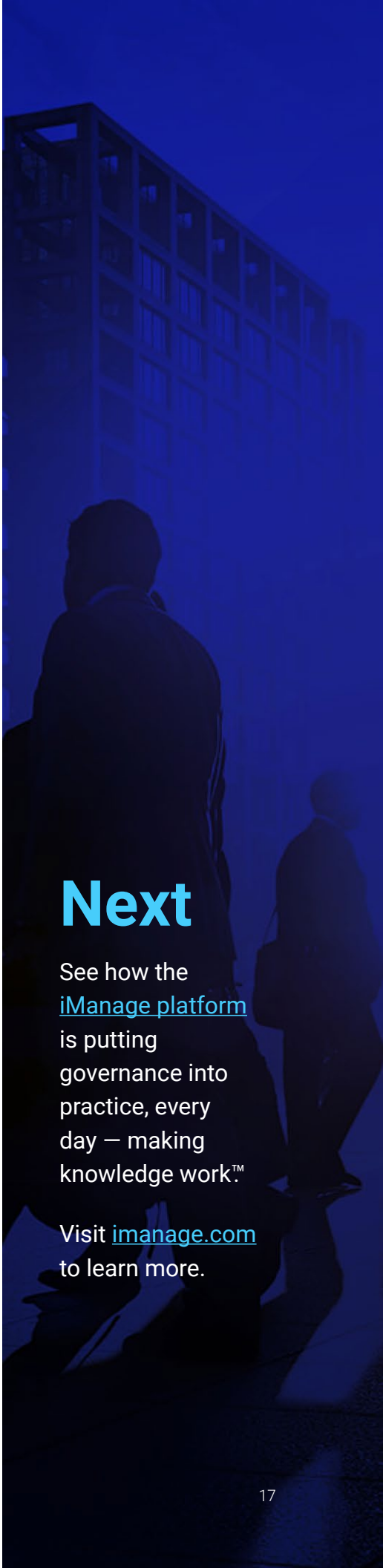
Governance is not the price an organization pays to use AI. It's the investment that makes AI trustworthy enough to rely on. **Not the handbrake, the ignition.**

A lawyer who trusts an AI draft the way they'd trust a colleague's judgment is making a similar mistake at an individual level that Air Canada made at a company level: its chatbot gave a confident, unwarranted answer, and the company was held responsible. Both cases trace back to a related failure: judgment that used to live in a person's head was never written down, never memorialized, and an agent had no way to inherit it.

AI makes producing a first draft easy. What's always been hard is knowing whether that output is true, relevant, and safe to act on. This judgment belongs to a person, not a system. Human expertise is what keeps the underlying data worth reasoning over in the first place. Even as more work is delegated to AI over time, the human in the loop is relied upon to preserve the integrity of ever-growing outputs.

iManage provides the foundation for agentic work: governed, well-disposed data, and a trail back to where an output came from, so decisions can scale without the trust behind them eroding the context an answer needs to be trustworthy. Preserved as it moves, so decisions can scale without losing the ability to withstand scrutiny.

As agentic AI becomes tightly woven into our business processes, governance is the common thread running through everything that follows. **Don't lose the thread.**



Next

See how the iManage platform is putting governance into practice, every day — making knowledge work.™

Visit imanage.com to learn more.

Sources

1. Statista, worldwide data creation projections, 2025–2026.
2. IDC FutureScape 2026, Worldwide IT Industry, Prediction 2.
3. Kiteworks, 2025 AI Data Security and Compliance Risk Study.
4. IBM, Cost of a Data Breach Report 2025, The AI Oversight Gap.
5. PagerDuty, Shadow AI Report, June 2026.
6. 2026 AI and Data Leadership Executive Benchmark Survey.
7. KPMG, Ten Key Regulatory Challenges, midyear 2026 outlook.
8. American Bar Association, ethics guidance on generative AI tools, 2024.
9. ConnectLive Chicago, The Agentic Future: Vision & Readiness, May 2026.
10. Civil Resolution Tribunal of British Columbia, Moffatt v. Air Canada, February 2024.

iManage is dedicated to Making Knowledge Work™

Our cloud-native platform is at the center of the knowledge economy, enabling every organization to work more productively, collaboratively, and securely. Built on more than 30 years of industry experience, iManage helps leading organizations manage documents and emails more efficiently, protect vital information assets, and leverage knowledge to drive better business outcomes. As your strategic business partner, we employ our award-winning AI-enabled technology, an extensive partner ecosystem, and a customer-centric approach to provide support and guidance you can trust to make knowledge work for you. iManage is relied on by more than one million professionals at 4,000 organizations around the world. Visit **www.imanage.com** to learn more.

iManage
Making knowledge work